(51) Int. Cl. 2: H 04 B 1/66

(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT

DE 28 11 454 A 1

(11) Offenlegungsschrift 28 11 454

(21) Aktenzeichen: P 28 11 454.7

(22) Anmeldetag: 14. 3. 78

(43) Offenlegungstag: 20. 9. 79

(30) Unionspriorität:

(32) (33) (31) —

(54) Bezeichnung: Verfahren zur Verbesserung der Wiedergabequalität bandbegrenzt verfügbarer Sprache

(71) Anmelder: Heinrich-Hertz-Institut für Nachrichtentechnik Berlin GmbH, 1000 Berlin

(72) Erfinder: Höhne, Hans Dietrich, Dr.-Ing., 1000 Berlin

DE 28 11 454 A 1

HEINRICH-HERTZ-INSTITUT FÜR NACHRICHTENTECHNIK BERLIN GMBH

Patentansprüche

1. Verfahren zur Verbesserung der Wiedergabequalität bandbegrenzt verfügbarer Sprache unter Verwendung von Zusatzinformation, die mit Hilfe des verfügbaren Signals bestimmt wird, gekennzeichnet durch folgende Verfahrensabschnitte:

   - aus Mustern des verfügbaren Sprachsignals werden Parameter X gewonnen;
   - von diesen Parametern $\underline{X}$ werden Abstände α_k zu Parametern $\underline{A}_k$ bestimmt - die Parameter $\underline{A}_k$ liegen von jeweils einem eine Äquivalenzklasse von Lauten bandbegrenzter Sprache charakterisierenden Prototyp abgespeichert vor;
   - den Parametern $\underline{A}_k$ entsprechende Parameter $\underline{B}_k$ werden abgerufen - auch die Parameter $\underline{B}_k$ liegen von jeweils einem eine Äquivalenzklasse von Lauten charakterisierenden Prototyp abgespeichert vor, jedoch von Sprache mit dem für die vorgesehene Wiedergabe erforderlichen Spektrum;
   - unter Berücksichtigung der Abstände α_k zwischen den Parametern $\underline{X}$ und den Parametern $\underline{A}_k$ werden die Parameter $\underline{B}_k$ zur Bildung der im verfügbaren Sprachsignal fehlenden spektralen Information herangezogen.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß bei der Gewinnung der Parameter $\underline{X}$ ein zusätzliches Fehlersignal gebildet und dieses Fehlersignal bei der Bildung der im verfügbaren Sprachsignal fehlenden spektralen Information mitherangezogen wird.

3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß abhängig von dem Verhältnis der Energien in Spektralbereichen des verfügbaren Sprachsignals die gebildete spektrale Information und das verfügbare Sprachsignal für die wiederzugebende Sprache zusammengefaßt werden.

4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß die Parameter $\underline{X}$, $\underline{A}_k$ und $\underline{B}_k$ der Sprachsignale Energien in spektralen Kanälen sind.

5. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß die Parameter $\underline{X}$, $\underline{A}_k$ und $\underline{B}_k$ der Sprachsignale Prädiktor- oder Reflexionskoeffizienten sind.

6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß die abgerufenen Parameter $\underline{B}_k$ mit zunehmenden Abständen α_k für die Bildung der im verfügbaren Sprachsignal fehlenden spektralen Information mit überproportional abnehmenden Anteilen herangezogen werden.

7. Verfahren nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, daß die Verbesserung der Wiedergabequalität der bandbegrenzt verfügbaren Sprache in Echtzeit erfolgt.

HEINRICH-HERTZ-INSTITUT FÜR NACHRICHTENTECHNIK BERLIN GMBH

Verfahren zur Verbesserung der Wiedergabequalität bandbegrenzt verfügbarer Sprache

Die Erfindung bezieht sich auf ein Verfahren zur Verbesserung der Wiedergabequalität bandbegrenzt verfügbarer Sprache unter Verwendung von Zusatzinformation, die mit Hilfe des verfügbaren Signals bestimmt wird. Mit fortschreitender technischer Entwicklung wachsen die Qualitätsanforderungen, so auch an Medien, mit denen übertragene Sprache wiedergegeben wird. Darüberhinaus ist es wirtschaftlich bedeutsam, wenn Bandbreite bei der Übertragung von Sprache ohne wesentlichen Qualitätsverlust bei der Wiedergabe eingespart werden kann, weil sich bei gegebener Breite eines Übertragungsbandes die dort unterzubringende Kanalzahl entsprechend erhöhen läßt. In manchen Fällen, z.B. beim beweglichen Landfunk, liegt hierin eine vorteilhafte Möglichkeit für eine Kapazitätsausweitung.

Die Einsparung von Bandbreite ohne wesentliche Minderung der Wiedergabequalität wird allgemein durch jeweils gegensinnig wirkende Maßnahmen auf der Sende- und auf der Empfangsseite herbeigeführt. Dazu wird senderseitig die Redundanz reduziert und z.B. mit Vocoderverfahren, mit adaptiver Differenz-Puls-Code-Modulation (ADPCM), mit Subbandcodierung oder auch mit Modulationsverfahren im analogen Bereich gearbeitet. Voraussetzung hierbei ist der Zugriff auf den Sender, so daß derartige Verfahren - falls nicht ein ausgewähltes Verfahren in sehr großem Umfang Einführung findet - auf regional und/oder anwendungstechnisch eng begrenzte Gebiete beschränkt bleiben müssen.

Bei der der Erfindung zugrundeliegenden Aufgabenstellung wird davon ausgegangen, daß ein solcher Zugriff zur Sendeseite nicht besteht, die Verbesserung der Wiedergabequalität bandbegrenzt verfügbarer Sprache also allein empfangsseitig erfolgen muß. Das bedeutet, die Grenzen des Bandes beim empfangenen Signal können in weiten Bereichen variieren, die für die Verbesserung der Wiedergabequalität zu treffenden Maßnahmen also in mehr oder weniger großem Umfang erforderlich sein, um insgesamt einen möglichst geringen Verlust an Silbenverständlichkeit und auch an Natürlichkeit zu erzielen.

Ein Teil dieser Problemstellungen ist bekannt (RLE Progress Report Nr. 119 (MIT, 1977),Seiten 100, 101). Der dort angegebene Weg sieht vor, tiefpaßgefilterte Sprache zu verbessern, indem fehlende spektrale Information allein empfangsseitig wieder eingesetzt wird. Wenn nur der niedrigfrequente Teil des Signals verfügbar ist, soll es danach möglich sein, einen großen Teil des fehlenden höherfrequenten Anteils aus der verfügbaren spektralen Energie zu bestimmen und damit die natürliche Sprache zu rekonstruieren. Ein wesentlicher Vorbehalt besteht darin, daß dieses bekannte Verfahren insbesondere für stimmhafte Sprache befriedigend arbeitet, bei der diskrete Frequenzen und Amplituden von Formanten gut ausgebildet sind. Ein den Frequenzgang formendes Filter soll dazu mit Harmonischen der aus dem verfügbaren Signal gewonnenen Grundfrequenz angeregt werden. Das erhaltene Signal mag zwar im Langzeitspektrum einem nicht bandbegrenzten Signal entsprechen; da jedoch der Vokaltrakt für jeden Menschen individuell ist und sich zudem bei jedem Laut ändert, sind überzeugende Ergebnisse dann nicht zu erwarten, wenn höhere zu ergänzende Formanten laut- und sprecherunabhängig zuzufügen sind.

Das Verfahren gemäß der Erfindung ist durch folgende Verfahrensabschnitte gekennzeichnet:

- aus Mustern des verfügbaren Sprachsignals werden Parameter $\underline{X}$ gewonnen;
- von diesen Parametern $\underline{X}$ werden Abstände α_k zu Parametern $\underline{A}_k$ bestimmt - die Parameter $\underline{A}_k$ liegen von jeweils einem eine Äquivalenzklasse von Lauten bandbegrenzter Sprache charakterisierenden Prototyp abgespeichert vor;
- den Parametern $\underline{A}_k$ entsprechende Parameter $\underline{B}_k$ werden abgerufen - auch die Parameter $\underline{B}_k$ liegen von jeweils einem eine Äquivalenzklasse von Lauten charakterisierenden Prototyp abgespeichert vor, jedoch von Sprache mit dem für die vorgesehene Wiedergabe erforderlichen Spektrum;
- unter Berücksichtigung der Abstände α_k zwischen den Parametern $\underline{X}$ und den Paramtern $\underline{A}_k$ werden die Parameter $\underline{B}_k$ zur Bildung der im verfügbaren Sprachsignal fehlenden spektralen Information herangezogen.

Diese Verfahrensabschnitte können auch als ein Erkennungs- und ein Syntheseabschnitt angesehen werden, bei denen auf abgespeicherte Information zurückgegriffen wird. Die Speichertechniken, die in engem Zusammenhang mit der Art des Syntheseverfahrens stehen, insbesondere jedoch die Informationsinhalte sind nach folgenden Gesichtspunkte zu bestimmen.

Die Verwendung abgespeichert vorliegender Information, passend zum verfügbaren Sprachsignal, kommt mit einer Filterung gemäß der Langzeitstatistik des Sprachsignals vieler Sprecher nicht aus. Deshalb wird - ähnlich wie bei der Spracherkennung, obwohl bei der Erfindung kein Spracherkennungsproblem im eigentlichen Sinne vorliegt - für die charakteristischen Laute und Lautgruppen der Sprache eine Klassifizierung vorgenommen. Prototypen solcher Äquivalenzklassen lassen sich als Vektoren genügend genau festlegen, also speichern. An sich wären technisch unrealistisch viele Äquivalenzklassen vorzusehen, um bei der Erkennungsphase die jeweils zutreffenden abgespeicherten Parameter bestimmen zu können. Das ist jedoch nicht erforderlich, d.h. die Zahl der Äquivalenzklassen kann auf weniger als 20, evtl. weniger als 10, beschränkt bleiben, weil die für das erfindungsgemäße Verfahren kennzeichnende Abstandsbestimmung der Parameter des Sprachmusters von den abgespeicherten Parametern von Prototypen einer Zerlegung in Parameter-Komponenten gleichkommt bzw. als Erkennung resultierender abgespeicherter Parameter anzusehen ist. Sodann ergibt sich die Synthese vom Grundsatz her aus einer Assoziation aufgrund der Erkennung, bei der die verwendeten abgespeicherten Parameter durch eine feste Zuordnung zu den erkannten vorgegeben werden und die Qualität der wiederzugebenden Sprache verbessern, weil von ihnen Laute bzw. Lautgruppen charakterisiert werden, die das für die vorgesehene Wiedergabe erforderliche Spektrum besitzen.

Sowohl für den Abschnitt der Erkennung als auch den der Synthese ist eine einfache Minimum-Maximum-Entscheidung denkbar. Der technische Aufwand für eine entsprechende Schaltung ist verhältnismäßig gering, erfordert jedoch - wie oben bereits erwähnt - bei hohen Qualitätsanforderungen an die wiederzugebende Sprache eine große Zahl von Äquivalenzklassen und damit große Speicher. Bevorzugte Ausführungsformen der Erfindung beruhen dagegen auf einem Mischen der durch die Erkennung bestimmten Anteile, aus denen sich die zur Qualitätsverbesserung verwendete Zusatzinformation zusammensetzt. Diese Zusatzinformation kann sowohl bezüglich der Quantität ihrer Anteile als auch im Verhältnis zum Anteil des in die Wiedergabe einbezogenen ursprünglich verfügbaren Sprachsignals bestimmt werden. Ein Fehlersignal, das bei der Gewinnung der Parameter aus dem Muster des verfügbaren Sprachsignals gebildet wird, ermöglicht eine einfache und wirkungsvolle Synthese der Zusatzinformation.

Von ebenfalls wesentlicher Bedeutung für Ausführungsformen der Erfindung ist die Möglichkeit, den Anteil von Zusatzinformation in der wiederzugebenden Sprache in Abhängigkeit von der Qualität des verfügbaren Sprachsignals bestimmen zu können. Sofern nämlich im verfügbaren Sprachsignal bereits spektrale Anteile enthalten sind, die durch abgespeicherte Parameter von Äquivalenzklassen von Lauten bandbegrenzter Sprache nicht oder nicht genügend Berücksichtigung finden würden, kann abhängig vom Verhältnis der Energien des verfügbaren Sprachsignals die wiederzugebende Sprache zusammengesetzt werden.

Die Mischungsverhältnisse, mit denen die abgespeicherten Prototypen zur Bildung der Zusatzinformation herangezogen werden, richten sich nicht nur schlechthin nach den Abständen zwischen den in der Erkennungsphase miteinander verglichenen Parametern, es ist vorteilhaft, wenn mit wachsenden Abständen die zur Bildung der Zusatzinformation heranzuziehenden Anteile überproportional abnehmen. Diese Maßnahme wirkt sich qualitativ in Richtung einer Minimum-Maximum-Entscheidung aus, ohne jedoch wirklich eine solche Entscheidung zu sein.

Insbesondere im Hinblick auf neue Technologien elektronischer Bauelemente (VLSI = Very Large Scale Integration) sind die wirtschaftlich-technischen Randbedingungen für Ausführungsformen der Erfindung günstig. Das Verfahren zur Verbesserung der Wiedergabequalität bandbegrenzt verfügbarer Sprache kann dann nicht nur z.B. bei Rundfunksendern o.ä. erfolgen, bei denen über Telefon empfangene Sprache aufgenommen, in ihrer Qualität verbessert und sodann ausgesendet wird, es kann vor allem in Echtzeit und am Ort des Teilnehmers erfolgen.

Im Zusammenhang mit dem in der Zeichnung dargestellten Blockschaltbild wird schematisch der Ablauf des Verfahrens gemäß der Erfindung näher erläutert:

Das verfügbare bandbegrenzte Sprachsignal ist mit $s_1(t)$ bezeichnet. Es wird außer zu einem Addierer (10) zu einem Filter (1) geführt. Dort erfolgt eine Parameterabschätzung, wobei es sich bei diesen - und den noch folgenden - Parametern jeweils um Energien in spektralen Kanälen oder um Prädiktorkoeffizienten handelt. Diese Parameter $\underline{X}$, die aus Mustern des verfügbaren bandbegrenzten Signals $s_1(t)$ gewonnen wurden, werden in einem Abstandsbildner (2) mit Paramtern $\underline{A}_k$ verglichen, die aus einem Speicher (3a) über einen Datenbus (4) zugeführt werden. Für die einzelnen Abstände der Parameter $\underline{X}$ zu den Prototypen

von Äquivalenzklassen ergeben sich damit Werte α_k, von denen abhängig ein betreffender Multiplizierer (5) mit den aus einem Speicher (3b) abgerufenen Parametern $\underline{B}_k$, die bezüglich der Äquivalenzklassen, nicht jedoch hinsichtlich ihrer Vektorkomponenten übereinstimmen, die einzelnen Anteile für die an einem Addierer (6) passend gebildete Zusatzinformation bestimmt wird. In einem Synthetisator (7) wird aus dieser Zusatzinformation ein Analogsignal gebildet, das entweder (- nicht dargestellt -) direkt oder über einen Regelverstärker (9) zum oben bereits erwähnten Addierer (10) gelangt, an dessen Ausgang das in der Qualität verbesserte wiederzugebende Signal $s_2(t)$ vorliegt.

Ist ein Regelverstärker (9) vorgesehen, wird in einem Mittelwertbestimmer (8) vom Signal $s_1(t)$ z.B. das Verhältnis der Energien im "Restband" zur Gesamtenergie bestimmt und der Regelverstärker (9) entsprechend eingestellt. Unter "Restband" werden die spektralen Anteile verstanden, die nach der Dimensionierung der Äquivalenzklassen wiederzugebender und bandbegrenzter Sprache im ungünstigsten Fall zuzufügen sind.

Im Blockschaltbild ist außerdem eine Verbindung zwischen dem Fitler (1) und dem Synthetisator (7) eingezeichnet. Diese Verbindung dient zur Übertragung eines Fehlersignals, das zur Erzeugung der Zusatzinformation benötigt wird.

Begnügt man sich mit Zusatzinformation, die nicht völlig sprecherunabhängig ist, kann noch eine zusätzliche grobe Klassifizierung - männliche/weibliche Stimme - vorgesehen werden.

Handelt es sich bei den Parametern $\underline{X}$, $\underline{A}_k$ und $\underline{B}_k$ um Pakorkoeffizienten, werden diese mit dem Eintreffen des Signals $s_1(t)$ z.B. blockweise berechnet. Das kann sukzessiv aus den Korrelationskoeffizienten der Fehlerfolgen bei Vorwärts- und Rückwärtsprädiktion mit einem Prädiktionsfehlerfilter in Kaskadenform durchgeführt werden. Sodann werden die Abstände $\alpha_k = |\underline{\pi} - \underline{\pi}_k|$ als Ähnlichkeitsmaß des empfangenen Signals zu den vorhandenen Äquivalenzklassen ermittelt. Aus den $\underline{\pi}_k^*$ wird dann gemäß

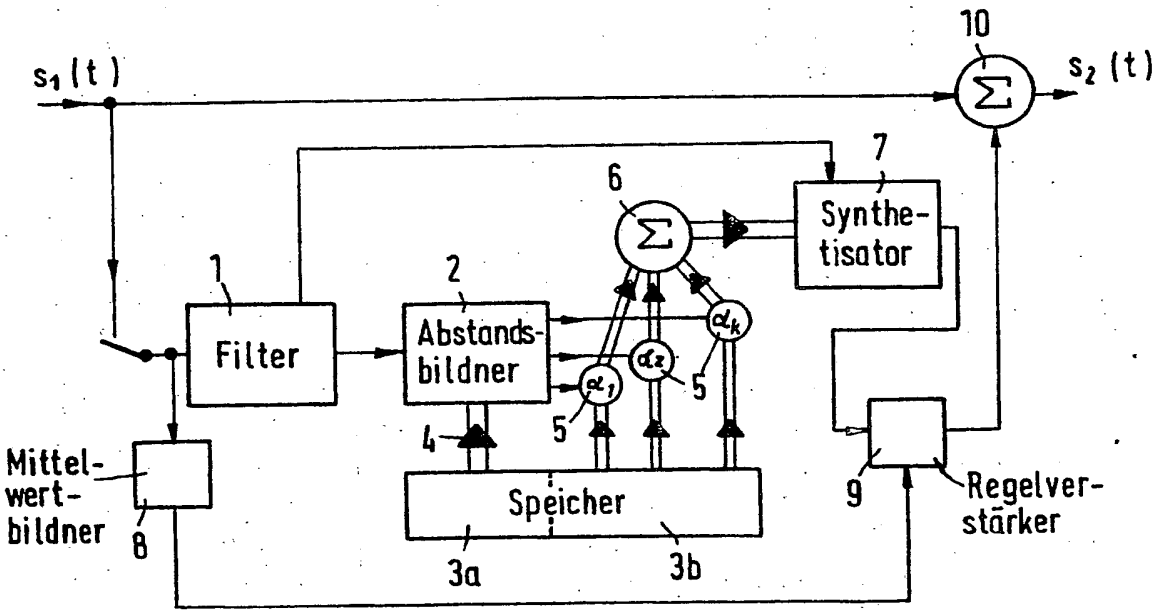
$$\underline{\pi}^* = \sum_{i=1}^{k} \alpha_i \, \underline{\pi}_i^*$$

ein Parkorkoeffizientensatz für das Restbandsignal erstellt, der dem Synthetisator zugeführt wird. Im Falle der Benützung eines Prädiktionsverfahrens empfiehlt sich die Verwendung der Parkorkoeffizienten, weil sich hierbei die Stabilität der Synthese leicht gewährleisten läßt. Der Synthetisator wird ebenfalls als Filter in Kaskadenform aufgebaut. Die Anregung erfolgt mit dem im Prädiktionsfehlerfilter gewonnenen Fehlersignal. Dieses Beispiel zeigt eine Verwertung der bei der Erkennung gewonnenen Abstandsmaße zur Berechnung der Parameter für die Synthese von Zusatzinformation.

Bei einer Parametergewinnung im Frequenzbereich wird das Sprachsignal durch Bandpässe in z.B. 10 Unterbänder aufgeteilt und die Energie in diesen Kanälen wird nach Gleichrichten und weiterer Tiefpaßtfilterung z.B. mit 25 Hz als Parameter betrachtet. Als Fehlersignal ist (wie beim Voice-Excited-Vocoder) das Basisband z.B. bis ca. 1000 Hz verwendbar. Zur Synthese werden Bandpaßfilter mit dem Restbandsignal angeregt und nach Spitzenbegrenzung zur Vermeidung von Amplitudenschwankungen mit den Vocoder-Kanal-Signalen moduliert. Der Unterschied zum reinen Voice-Excited-Vocoder besteht darin, daß die Vocoder-Kanal-Signale nicht übertragen werden, sondern als Parameter im folgenden Mustererkennungsprozeß dienen. Dabei werden wiederum Abstände $\alpha_1 \ldots \alpha_k$ zu gespeicherten Parametern für Lautprototypen in bandbegrenztem Signal berechnet und daraus und aus gespeicherten Prototypen des breitbandigen oder des Restbandsignals neue Vocoder-Kanal-Signale entwickelt.

Nummer: 28 11 454
Int. Cl.²: H 04 B 1/66
Anmeldetag: 14. März 1978
Offenlegungstag: 20. September 1979

2811454



909838/0392